

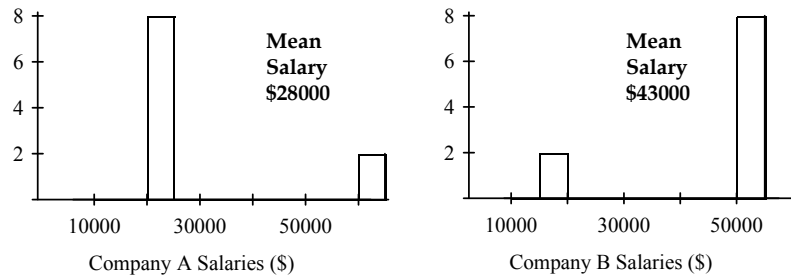
Chapter 2

Displaying and Describing Categorical Data

Solutions to Class Examples:

1. See Class Example 1.
2. Answers will vary according to your data.
3. See Class Example 3.
4. Point out that this example deals with quantitative variables, not categorical. The paradox can be explained when you realize that Company A must employ a greater percentage of laborers than Company B. Also, Company A must employ a smaller percentage of managers than Company B. If laborers earn salaries that are considerably lower than managers, the salaries of Company A's laborers will pull the company average down, and the salaries of Company B's managers will pull the company average up.

The proper way to compare the companies is to use the salaries that are broken down by job type. Using the overall average salary leads to a misleading conclusion.



Investigative Task

Race and the Death Penalty uses a three-way contingency table and requires comparing marginal and conditional distributions.

Supplemental Resources

After the Investigative Task, there is a worksheet on the relationship between smoking and education level.

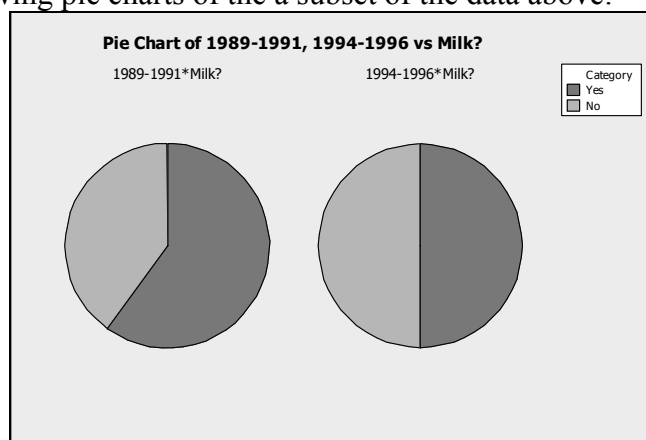
Statistics Quiz A – Chapter 2

Name _____

Has the percentage of young girls drinking milk changed over time? The following table is consistent with the results from “Beverage Choices of Young Females: Changes and Impact on Nutrient Intakes” (Shanthy A. Bowman, *Journal of the American Dietetic Association*, 102(9), pp. 1234-1239):

		Nationwide Food Survey Years			
		1987-1988	1989-1991	1994-1996	Total
Drinks Fluid Milk	Yes	354	502	366	1222
	No	226	335	366	927
	Total	580	837	732	2149

- Find the following:
 - What percent of the young girls reported that they drink milk? _____
 - What percent of the young girls were in the 1989-1991 survey? _____
 - What percent of the young girls who reported that they drink milk were in the 1989-1991 survey? _____
 - What percent of the young girls in 1989-1991 reported they drink milk? _____
- What is the marginal distribution of milk consumption?
- Do you think that milk consumption by young girls is independent of the nationwide survey year? Use statistics to justify your reasoning.
- Consider the following pie charts of the a subset of the data above:



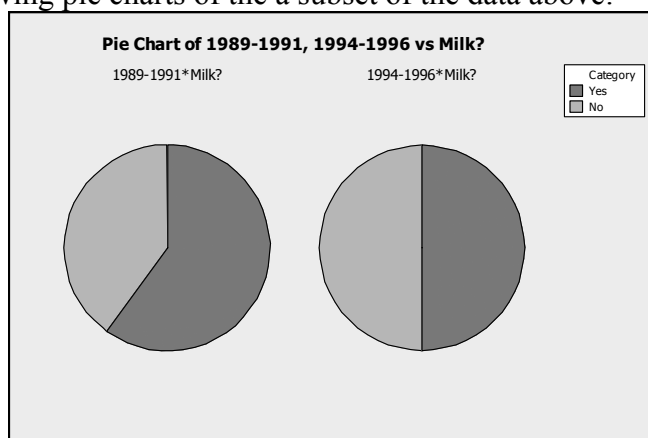
Do the pie charts above indicate that milk consumption by young girls is independent of the nationwide survey year? Explain.

Statistics Quiz A – Chapter 2 - Key

Has the percentage of young girls drinking milk changed over time? The following table is consistent with the results from “Beverage Choices of Young Females: Changes and Impact on Nutrient Intakes” (Shanthy A. Bowman, *Journal of the American Dietetic Association*, 102(9), pp. 1234-1239):

		Nationwide Food Survey Years			
		1987-1988	1989-1991	1994-1996	Total
Drinks Fluid Milk	Yes	354	502	366	1222
	No	226	335	366	927
	Total	580	837	732	2149

- Find the following:
 - What percent of the young girls reported that they drink milk? 56.9%
 - What percent of the young girls were in the 1989-1991 survey? 38.9%
 - What percent of the young girls who reported that they drink milk were in the 1989-1991 survey? 41.1%
 - What percent of the young girls in 1989-1991 reported they drink milk? 60.0%
- What is the marginal distribution of milk consumption?
Yes: 56.9%; No: 43.1%
- Do you think that milk consumption by young girls is independent of the nationwide survey year? Use statistics to justify your reasoning.
No. 56.9% of all young girls surveyed reported drinking milk, but 60% of the young girls reported drinking milk in the 1989-1991 survey. Since these percentages differ, milk consumption and year are not independent.
- Consider the following pie charts of the a subset of the data above:



Do the pie charts above indicate that milk consumption by young girls is independent of the nationwide survey year? Explain.

No. It looks like there is some sort of relationship between milk consumption and nationwide survey year, since the percentage of young girls who reported drinking milk is a larger slice of the pie chart for the 1989-1991 survey than the same response for the 1994-1996 survey.

Statistics Quiz B – Chapter 2**Name** _____

To determine if people's preference in dogs had changed in the recent years, organizers of a local dog show asked people who attended the show to indicate which breed was their favorite. This information was compiled by dog breed and gender of the people who responded. The table summarizes the responses.

1. Identify the variables and tell whether each is categorical or quantitative.

	Female	Male	Total
Yorkshire Terrier	73	59	132
Dachshund	49	47	96
Golden Retriever	58	33	91
Labrador	37	41	78
Dalmatian	45	28	73
Other breeds	86	67	153
Total	348	275	623

2. Which of the W's are unknown for these data?

3. Find each percent.

- a. What percent of the responses were from males who favor Labradors? _____
- b. What percent of the male responses favor Labradors? _____
- c. What percent of the people who choose Labradors were males? _____

4. What is the marginal distribution of breeds?

5. Write a sentence or two about the conditional relative frequency distribution of the breeds among female respondents.

6. Do you think the breed selection is independent of gender? Give statistical evidence to support your conclusion.

Statistics Quiz B – Chapter 2 – Key

To determine if people's preference in dogs had changed in the recent years, organizers of a local dog show asked people who attended the show to indicate which breed was their favorite. This information was compiled by dog breed and gender of the people who responded. The table summarizes the responses.

1. Identify the variables and tell whether each is categorical or quantitative.

Gender and Breed; both categorical.

	Female	Male	Total
Yorkshire Terrier	73	59	132
Dachshund	49	47	96
Golden Retriever	58	33	91
Labrador	37	41	78
Dalmatian	45	28	73
Other breeds	86	67	153
Total	348	275	623

2. Which of the W's are unknown for these data?

We do not know how or when the people were surveyed, or where the local dog show was located.

3. Find each percent.

- a. What percent of the responses were from males who favor Labradors? 6.6%
- b. What percent of the male responses favor Labradors? 14.9%
- c. What percent of the people who choose Labradors were males? 52.6%

4. What is the marginal distribution of breeds?

There were 132 Yorkshire terrier responses, 96 Dachshund responses, 91 Golden Retriever responses, 78 Labrador responses, 73 Dalmatian responses, and 153 Other responses.

5. Write a sentence or two about the conditional relative frequency distribution of the breeds among female respondents.

Among females, 20.9% chose Yorkshire Terriers, 14.2% Dachshunds, 16.7% Golden Retrievers, 10.6% Labs, and 12.9% Dalmatians. The remaining 24.7% of females preferred other breeds.

6. Do you think the breed selection is independent of gender? Give statistical evidence to support your conclusion.

The breed selection does not appear to be independent of gender. Overall, 56% of the respondents were females, but females were over-represented among those who favored Golden Retrievers (64%) and Dalmatians (62%), yet a much lower percentage (47%) among those who chose Labradors.

Statistics Quiz C – Chapter 2**Name** _____

In order to plan transportation and parking needs at a private high school, administrators asked students how they get to school. Some rode a school bus, some rode in with parents or friends, and others used “personal” transportation – bikes, skateboards, or just walked. The table summarizes the responses from boys and girls.

	Male	Female	Total
Bus	30	34	64
Ride	37	45	82
Personal	19	23	42
Total	86	102	188

- Identify the variables and tell whether each is categorical or quantitative.
- Which of the W’s are unknown for these data?
- Find each percent.
 - What percent of the students are girls who ride the bus? _____
 - What percent of the girls ride the bus? _____
 - What percent of the bus riders are girls? _____
- What is the marginal distribution of gender?
- Write a sentence or two about the conditional relative frequency distribution of modes of transportation for the boys.
- Do you think mode of transportation is independent of gender? Give statistical evidence to support your conclusion.

Statistics Quiz C – Chapter 2 – Key

In order to plan transportation and parking needs at a private high school, administrators asked students how they get to school. Some rode a school bus, some rode in with parents or friends, and others used “personal” transportation – bikes, skateboards, or just walked. The table summarizes the responses from boys and girls.

	Male	Female	Total
Bus	30	34	64
Ride	37	45	82
Personal	19	23	42
Total	86	102	188

1. Identify the variables and tell whether each is categorical or quantitative.

Gender and mode of transportation, both categorical.

2. Which of the W’s are unknown for these data?

We don’t know how or when the students were surveyed, nor where the school is.

3. Find each percent.

- a) What percent of the students are girls who ride the bus? 18.1%
- b) What percent of the girls ride the bus? 33.3%
- c) What percent of the bus riders are girls? 53.1%

4. What is the marginal distribution of gender?

There are 86 males and 102 females.

5. Write a sentence or two about the conditional relative frequency distribution of modes of transportation for the boys.

More boys (43%) caught rides to school than any other means of transportation. 35% rode the bus while only 22% used personal transportation like biking, skateboarding, or walking.

6. Do you think mode of transportation is independent of gender? Give statistical evidence to support your conclusion.

The way students get to school does seem to be independent of gender. Overall, 34% of students ride the bus, compared to 35% of the boys and 33% of the girls. 44% of all students caught rides with someone and 22% used personal transportation, almost the same as the percentages for boys (43% and 22%) or girls (44% and 23%) separately. These data provide little indication of a difference in mode of transportation between boys and girls at this school.

Race and the Death Penalty

In 1976 the Supreme Court ruled that the death penalty does not violate the U.S. Constitution's ban on "cruel and unusual punishments." Since then many states have passed capital punishment statutes, and over 500 convicted murderers have been executed nationwide.

Capital punishment may be constitutional, but there continues to be a debate about whether or not it is fair. One of the major issues in this debate involves race – the race of both the defendant and the murder victim. The central question: is justice blind?

In 1998 the Death Penalty Information Center published *The Death Penalty in Black and White*, a study examining the sentences following 667 murder convictions in Philadelphia courts between 1983 and 1993. This 3-way table shows how many death sentences were given among all the murder convictions.

DEATH SENTENCES	Black Victim	White Victim	Total
Black Defendant	76 of 422	21 of 99	97 of 521
White Defendant	1 of 25	17 of 121	18 of 146
Total	77 of 447	38 of 220	115 of 667

Is our system of justice colorblind in the administration of the death penalty? Based upon the above information, write a newspaper article discussing the association between race and death sentences in the United States.

(Don't forget: the best analyses of data usually combine visual, numerical, and verbal descriptions.)

Statistics Task**Chapter 2**

	Components	Comments
<i>Think</i>	1. Identifies useful marginal and conditional distributions (or %'s) to make effective comparisons	
<i>Show</i>	2. ○ Numerical – no major errors ○ Visual – includes comparative pie charts or (segmented?) bar graphs ○ well-labeled & fairly accurate	
<i>Tell</i>	3. Verbal - Written article ... ○ is clear and concise ○ identifies the W's ○ is in the proper context ○ uses vocabulary correctly ○ avoids speculation 4. States a conclusion about the association between race and the death penalty, explaining at least two examples of statistical support	

Components are scored as **Essentially correct**, **Partially correct**, or **Incorrect**

1: Use of marginal and conditional distributions

E - Demonstrates understanding of marginal vs conditional for comparison

P - Calculates useful %'s but may not understand why, or has major arithmetic errors

I – Calculations are not %'s, or not useful for comparisons

2: Graphical display

E – Uses comparative pie/bar graph, well-labeled and fairly accurate

P – Graph is comparative but poorly constructed or explained

I – Graph is not comparative or is missing

3: News article

E – Article has all 5 listed properties

P – Article has 3 or 4 of the listed properties

I – Article has fewer than 3 of the properties

4: Conclusion

E – Correct general conclusion is supported by 2 appropriate comparisons

P – Conclusion is not clearly stated, or only one supporting comparison is given

I – Conclusion is incorrect, unsupported by statistics, or missing.

Scoring

- E's count 1 point, P's are 1/2
- Grade: A = 4, B = 3, etc., with +/- based on rounding (ex: 3.5 rounded to 3 is a B+)

Name _____

Grade _____

NOTE: We present a model solution with some trepidation. This is not a scoring key, just an example. Many other approaches could fully satisfy the requirements outlined in the scoring rubric. That (not this) is the standard by which student responses should be evaluated.

Model Solution – Investigative Task – Race and the Death Penalty

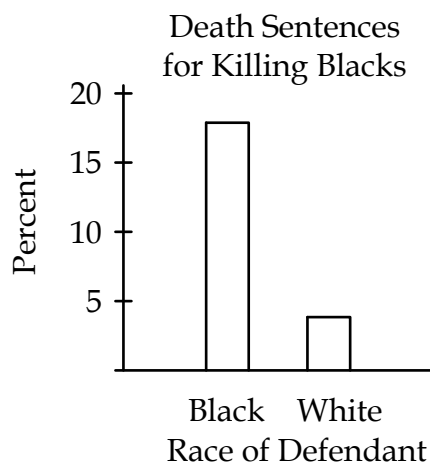
In the 10 years between 1983 and 1993, the city of Philadelphia saw 667 defendants convicted of murder. Of these, 17% were sentenced to death. Was the death penalty administered without regard to the race of the defendant or the victim? Is justice colorblind?

Overall, blacks convicted of murder were sentenced to the death penalty in 18.6% of cases. The death penalty was the sentence for only 12.3% of whites convicted of murder. There may be evidence that black defendants fared worse in regard to death penalty sentences.

At first, there doesn't appear to be an association between the race of the victim and the rate of death sentences. Defendants convicted of murdering whites were sentenced to death about 17% of the time, as were defendants convicted of murdering blacks. But these percentages are misleading. It's not until we look a bit deeper that we see the true picture.

When race of the defendant and race of the victim are both taken into account, blacks are convicted at higher rates across the board than whites. When the victim was black, 18% of black defendants were sentenced to death, compared to only 4% for white defendants. Likewise, when the victim was white, blacks were sentenced to death 21% of the time, while only 14% of white defendants were sentenced to death. Note also that, no matter what the race of the defendant, death penalty rates for killing whites are higher than the rates for killing blacks, 21% to 18% for black defendants, and 14% to 4% for white defendants.

These startling statistics present evidence that justice is not blind to color, at least not in the city of Philadelphia.



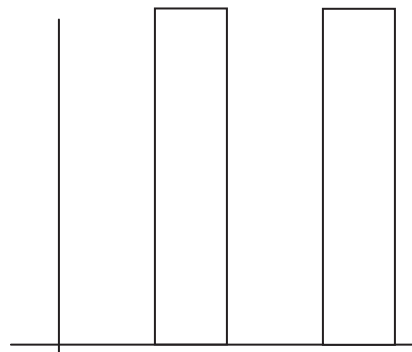
*Statistics**Chapter 2****Smoking and Education***

200 adults shopping at a supermarket were asked about the highest level of education they had completed and whether or not they smoke cigarettes. Results are summarized in the table.

	Smoker	Non-smoker	Total
High school	32	61	93
2 yr college	5	17	22
4+ yr college	13	72	85
Total	50	150	200

1. Discuss the W's.
2. Identify the variables.
3. a) What percent of the shoppers were smokers with only high school educations? _____
 b) What percent of the shoppers with only high school educations were smokers? _____
 c) What percent of the smokers had only high school educations? _____

4. Create a segmented bar graph comparing education level among smokers and non-smokers. Label your graph clearly



5. Do these data suggest there is an association between smoking and education level? Give statistical evidence to support your conclusion.
6. Follow-up question: Does this indicate that students who start smoking while in high school tend to give up the habit if they complete college? Explain.

Statistics – Smoking and Education Key

- Who: 200 adults
What: education level and smoking habits
When: not specified
Where: shopping mall
How: not specified. Was this a random sample, or were some people simply asked?
Why: to examine possible links between smoking and education level
- Categorical variables: Education level, and whether or not the person was a smoker.
- $\frac{32}{200} = 16\%$
 - $\frac{32}{93} \approx 34.4\%$
 - $\frac{32}{50} = 64\%$
- The segmented bar graph comparing education level among smokers and non-smokers is at the right.
- These data provide evidence of an association between smoking and education level. 64% of smokers had only a high school diploma, while only 40.7% of non-smoker had only high school diplomas. Only 26% of smokers had four or more years of college, compared to 48% of smokers.
- These data do not indicate that students who start smoking in high school tend to give up the habit if they complete college. These data were gathered at one time, about two different groups, smokers and non-smokers. We have no idea if smoking behavior changes over time.

